

УТВЕРЖДАЮ
Первый проректор, проректор
по УР

_____ А.Е. Рудин

Рабочая программа дисциплины

Б1.В.ДВ.03.01 Технологии анализа больших данных

Учебный план: 2025-2026 09.03.03 ИИТА ИТ-решения и ОБП ОО №1-1-165.plx

Кафедра: **33** Цифровых и аддитивных технологий

Направление подготовки:
(специальность) 09.03.03 Прикладная информатика

Профиль подготовки: ИТ-решения и обеспечение бизнес-процессов
(специализация)

Уровень образования: бакалавриат

Форма обучения: очная

План учебного процесса

Семестр (курс для ЗАО)	Контактная работа обучающихся		Сам. работа	Контроль, час.	Трудоё мкость, ЗЕТ	Форма промежуточной аттестации	
	Лекции	Практ. занятия					
7	УП	16	16	49	27	3	Экзамен
	РПД	16	16	49	27	3	
Итого	УП	16	16	49	27	3	
	РПД	16	16	49	27	3	

Рабочая программа дисциплины составлена в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 09.03.03 Прикладная информатика, утвержденным приказом Минобрнауки России от 19.09.2017 г. № 922

Составитель (и):

Доцент

Лебедева Светлана
Викторовна

От кафедры составителя:

Заведующий кафедрой цифровых и аддитивных технологий

Сошников Антон
Владимирович

От выпускающей кафедры:

Заведующий кафедрой

Сошников Антон
Владимирович

Методический отдел:

1 ВВЕДЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

1.1 Цель дисциплины: Сформировать компетенции обучающегося в области теоретических и практических основ технологий анализа больших данных, получение знаний об особенностях анализа данных для решения разнообразных бизнес-аналитических задач, возникающих в профессиональной деятельности.

1.2 Задачи дисциплины:

Освоение методологии обработки больших данных;
 Владение методами и технологиями работы с данными и знаниями;
 Иметь представление о структуризации данных и технологиях Knowledge Discovery in Databases и Data Mining;
 Освоение основ статистического анализа данных.

1.3 Требования к предварительной подготовке обучающегося:

Предварительная подготовка предполагает создание основы для формирования компетенций, указанных в п. 2, при изучении дисциплин:

- Операционные системы, сети и телекоммуникации
- Теория систем и системный анализ
- Прикладная статистика
- Алгоритмизация и программирование
- Программная инженерия
- Исследование операций и методы оптимизации
- Облачные и виртуальные технологии
- Аналитика бизнес-процессов

2 КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-1: Способен разрабатывать процедуры интеграции программных модулей.
Знать: Основные понятия и терминологию в области технологий обработки и анализа больших данных; языки программирования для обработки данных и методы организации взаимодействия аналитических платформ с распределенными хранилищами информации.
Уметь: Разрабатывать программный код для эффективной обработки распределенных данных большого объема с использованием языков программирования; использовать программные средства визуализации и интерактивного исследования больших данных.
Владеть: Программными средствами автоматической загрузки и обработки больших данных; навыками работы с системой интерактивного исследования и визуализации больших данных.
ПК-8: Способен организовывать оценку соответствия требований существующих систем и их аналогов.
Знать: Основные законы естественнонаучных дисциплин при тестировании программного обеспечения с учетом обработки больших объемов данных.
Уметь: Строить управляющий граф программы для тестирования, на основании использования основных законов естественнонаучных дисциплин, и современные информационно - коммуникационные технологии в профессиональной деятельности.
Владеть: Навыками разработки эффективных наборов тестов для простых и крупных информационных систем.

3 РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Наименование и содержание разделов, тем и учебных занятий	Семестр (курс для ЗАО)	Контактная работа		СР (часы)	Инновац. формы занятий	Форма текущего контроля
		Лек. (часы)	Пр. (часы)			
Раздел 1. Технологии анализа данных	7					О
Тема 1. Введение в анализ больших данных. Методология анализа данных. Практические занятия: Аналитический и информационный подходы к моделированию.		2	2	5	ИЛ	
Тема 2. Инженерия данных. Эволюция инженерии данных. Инженерия данных и наука о данных. Инженерия жизненного цикла данных. Основные фоновые процессы. Управление данными. Практические занятия: Этапы моделирования, формы представления данных. Подготовка данных к анализу		2	2	8	ИЛ	

Раздел 2. Технологии анализа больших объемов данных					
Тема 3. BIG DATA Ключевые тренды Big Data. Главные характеристики Big Data. Фазы жизненного цикла. Практические занятия: Подготовка данных к анализу. Очистка данных.		2	2	8	ИЛ
Тема 4. Технологии аналитики больших данных. Data Mining, Машинное обучение, Business Intelligence. Основные понятия. Жизненный цикл проекта Data Science. Дескриптивная аналитика. Диагностическая аналитика. Предиктивная аналитика. Прескриптивная аналитика. Принципы анализа данных. Практические занятия: Подготовка данных к анализу. Преобразование данных. Извлечение и визуализация данных.		4	4	10	ИЛ
Тема 5. Современные технологии хранения данных. DWH, Data Lake и Data Swamp. Современные платформенные решения хранения и анализа больших данных. Практические занятия: Разделение исходного набора данных на таблицы фактов и таблицы измерений.		2	2	8	ИЛ
Тема 6. Методы многомерного моделирования. Подходы Ральфа Кимбалла и Билла Инмона. Архитектура хранилищ данных. Архитектура Lakehouse. Модели данных хранилища данных. Модель для витрин данных. Модель на основе денормализованного подхода. Практические занятия: разработка модели хранилища данных. Извлечение и визуализация данных.		4	4	10	ИЛ
Итого в семестре (на курсе для ЗАО)		16	16	49	
Консультации и промежуточная аттестация (Экзамен)			2,5	24,5	
Всего контактная работа и СР по дисциплине			34,5	73,5	

4 КУРСОВОЕ ПРОЕКТИРОВАНИЕ

Курсовое проектирование учебным планом не предусмотрено

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

5.1 Описание показателей, критериев и системы оценивания результатов обучения

5.1.1 Показатели оценивания

Код компетенции	Показатели оценивания результатов обучения	Наименование оценочного средства
ПК-1	Формулирует основные понятия и терминологию в области больших данных.	Вопросы устного собеседования
	Классифицирует языки программирования для обработки больших объемов данных с использованием средств визуализации и интерактивного исследования.	Практико-ориентированные задания
	Демонстрирует программный код для организации автоматической загрузки и обработки больших данных.	Практико-ориентированные задания

ПК-8	Раскрывает сущность основных естественнонаучных законов используемых при тестировании программного обеспечения с учетом обработки больших объемов данных.	Вопросы устного собеседования
	Строит алгоритм программы для тестирования, на основании использования основных законов естественнонаучных дисциплин.	Практико-ориентированные задания
	Демонстрирует тест для тестирования информационной системы.	Практико-ориентированные задания

5.1.2 Система и критерии оценивания

Шкала оценивания	Критерии оценивания сформированности компетенций	
	Устное собеседование	Письменная работа
5 (отлично)	Полный, исчерпывающий ответ, явно демонстрирующий глубокое понимание предмета и широкую эрудицию в оцениваемой области. Критический, оригинальный подход к материалу. Качество исполнения всех элементов практико-ориентированного задания полностью соответствует всем требованиям Учитываются баллы, накопленные в течение семестра.	.
4 (хорошо)	Ответ полный, основанный на проработке всех обязательных источников информации. Подход к материалу ответственный, но стандартный. Практико-ориентированное задание выполнено в достаточном объеме, но ограничивается только основными подходами. Учитываются баллы, накопленные в течение семестра.	
3 (удовлетворительно)	Ответ стандартный, в целом качественный, основан на всех обязательных источниках информации. Присутствуют небольшие пробелы в знаниях или несущественные ошибки. Практико-ориентированное задание выполнено в соответствии с заданием.	
	Имеются отдельные несущественные ошибки или отступления от правил оформления работы. Учитываются баллы, накопленные в течение семестра	
2 (неудовлетворительно)	Неспособность ответить на вопрос без помощи экзаменатора. Незнание значительной части принципиально важных элементов дисциплины. Многочисленные грубые ошибки. Отсутствие одного или нескольких обязательных элементов практико-ориентированного задания, либо многочисленные грубые ошибки в работе, либо грубое нарушение правил оформления или сроков представления работы. Не учитываются баллы, накопленные в течение семестра.	

5.2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

5.2.1 Перечень контрольных вопросов

№ п/п	Формулировки вопросов
Семестр 7	
1	Big Data, определение, предпосылки, эволюция развития.
2	Big Data, определение, источники и виды данных.
3	Инженерия данных, определение, эволюция развития основных технологий .
4	Инженерия данных и наука о данных, сравнительный анализ.
5	Инженерия жизненного цикла данных. Генерация, хранение, поглощение.
6	Инженерия данных. Аналитика данных.
7	Инженерия данных. Обратный ETL.
8	Инженерия данных. Основные фоновые процессы.
9	Применение Big Data, основные технологические тренды.
10	Основные характеристики Big Data. 4 V.
11	Основные характеристики Big Data. 7 V.
12	Современные технологии управления большими данными. MapReduce и Hadoop.
13	Технологии аналитики больших данных. Методология OSEMN.
14	Технологии аналитики больших данных. Дескриптивная аналитика
15	Технологии аналитики больших данных. Предиктивная аналитика.
16	Технологии аналитики больших данных. Прескриптивная аналитика
17	Технологии аналитики больших данных. Диагностическая аналитика.
18	Современные технологии хранения данных. Data Lake и Data Swamp.
19	Современные технологии хранения данных. Lakehouse.
20	Методы многомерного моделирования. Модель для витрин данных.
21	Методы многомерного моделирования. Модель на основе денормализованного подхода.
22	Современные технологии хранения данных. DWH.

5.2.2 Типовые тестовые задания

Не предусмотрено.

5.2.3 Типовые практико-ориентированные задания (задачи, кейсы)

Типовые тестовые задания находятся в Приложении к данной РГД

5.3 Методические материалы, определяющие процедуры оценивания знаний, умений, владений (навыков и (или) практического опыта деятельности)

5.3.1 Условия допуска обучающегося к промежуточной аттестации и порядок ликвидации академической задолженности

Проведение промежуточной аттестации регламентировано локальным нормативным актом СПбГУПТД «Положение о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся»

5.3.2 Форма проведения промежуточной аттестации по дисциплине

Устная Письменная Компьютерное тестирование Иная

5.3.3 Особенности проведения промежуточной аттестации по дисциплине

- время на подготовку к устному собеседованию составляет 30 минут;
- выполнение кейс-задания осуществляется на компьютере за 60 минут.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Учебная литература

Автор	Заглавие	Издательство	Год издания	Ссылка
6.1.1 Основная учебная литература				
Мельниченко, А. С.	Математическая статистика и анализ данных	Москва: Издательский Дом МИСиС	2018	http://www.iprbookshop.ru/78563.html
Пальмов, С. В.	Интеллектуальный анализ данных	Самара: Поволжский государственный университет телекоммуникаций и информатики	2017	http://www.iprbookshop.ru/75376.html
6.1.2 Дополнительная учебная литература				

Адлер, Ю. П., Черных, Е. А.	Статистическое управление процессами. «Большие данные»	Москва: Издательский Дом МИСиС	2016	http://www.iprbookshop.ru/64199.html
Жуковский, О. И.	Информационные технологии и анализ данных	Томск: Томский государственный университет систем управления и радиоэлектроники, Эль Контент	2014	http://www.iprbookshop.ru/72106.html
Шифф В. К.	Корреляционный, регрессионный и дисперсионный анализ	СПб.: СПбГУПТД	2019	http://publish.sutd.ru/tp_ext_inf_publish.php?id=201950

6.2 Перечень профессиональных баз данных и информационно-справочных систем

Информационная система «Единое окно доступа к образовательным ресурсам. Раздел. Информатика и информационные технологии» [Электронный ресурс]. URL: http://window.edu.ru/catalog/?p_rubr=2.2.75.6
 Электронно-библиотечная система IPRbooks [Электронный ресурс]. URL: <http://www.iprbookshop.ru/>

6.3 Перечень лицензионного и свободно распространяемого программного обеспечения

MicrosoftOfficeProfessional
 Microsoft Windows
 Deductor Academic

6.4 Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Аудитория	Оснащение
Лекционная аудитория	Мультимедийное оборудование, специализированная мебель, доска
Компьютерный класс	Мультимедийное оборудование, компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду

Приложение

рабочей программы дисциплины Технологии анализа больших данных

по направлению подготовки 09.03.03 Прикладная информатика

наименование ОП (профиля): ИТ-решения и обеспечение аналитических бизнес-процессов

5.2.3 Типовые практико-ориентированные задания (задачи, кейсы)

№ п/п	Условия типовых практико-ориентированных заданий (задач, кейсов)		
1	Провести регрессионный анализ зависимости выручки от числа торговых точек.		Таблица 1
	Исходные данные		
	№	Выручка, тыс. руб.	Число торговых точек
	1	1598	2
	2	2644	5
	3	2197	4
	4	1959	5
	5	1052	3
	6	1922	3
	7	2385	5
	8	2581	5
	9	3105	5
	10	3896	4
	11	1510	4
	12	1880	2
	13	3620	5
	14	5002	6
	15	2819	5
2	Провести корреляционный анализ зависимости выручки от числа торговых точек.		Таблица 1
	Исходные данные		
	№	Выручка, тыс. руб.	Число торговых точек
	1	1598	2
	2	2644	5
	3	2197	4
	4	1959	5
	5	1052	3
	6	1922	3
	7	2385	5
	8	2581	5
	9	3105	5
	10	3896	4
	11	1510	4
	12	1880	2
	13	3620	5
	14	5002	6
	15	2819	5